

AD-A121 504

INTERPRETATION OF ELECTROANALYTICAL DATA USING
COMPUTERIZED PATTERN RECOGNITION(U) PURDUE UNIV
LAFAYETTE IN DEPT OF CHEMISTRY S P PERONE ET AL.
SEP 82 TR-29 N00014-75-C-0874

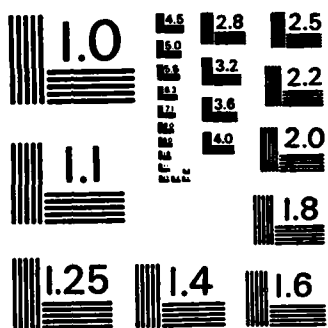
1/1

UNCLASSIFIED

F/G 9/2

NL

END
FILED
1
074



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A121504

12

OFFICE OF NAVAL RESEARCH

Contract No. N00014-75-C-0874

Task No. NR 051-552

TECHNICAL REPORT NO. 29

Interpretation of Electroanalytical Data
Using Computerized Pattern Recognition

by

S. P. Perone
Lawrence Livermore National Laboratory
Livermore, CA 94550
and Purdue University
W. Lafayette, IN 47907

Lars Kryger
Aarhus University
Denmark

W. A. Byers
Westinghouse Electric Corp.
Pittsburgh, PA 15235

Prepared for Publication in

Analytical Chemistry

September, 1982

DTIC
NOV 17 1982
A

Production in whole or in part is permitted for
any purpose of the United State Government.

Approved for Public Release; Distribution Unlimited

00

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 29	2. GOVT ACCESSION NO. AD A121 524	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Interpretation of Electroanalytical Data Using Computerized Pattern Recognition		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) S. P. Perone, L. Kryger, W. A. Byers		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0874
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Chemistry W. Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 051-552
11. CONTROLLING OFFICE NAME AND ADDRESS Materials Science Division Office of Naval Research Arlington, VA 22217		12. REPORT DATE September, 1982
		13. NUMBER OF PAGES 43
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release, Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Submitted for publication in <u>Analytical Chemistry</u>		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This article summarizes about a decade of work investigating the application of pattern recognition to the interpretation of electroanalytical data. Earliest work involved detection of unresolved doublet peaks in linear sweep voltammetry. These applications have now evolved to where on-line nearly real-time interpretation is possible to obtain electrode mechanism, kinetics, functional group, structural and activity information for organic compounds from a single voltammetric experiment. —		

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

BACKGROUND	1
PRINCIPLES OF PATTERN RECOGNITION	3
APPLICATIONS OF PATTERN RECOGNITION TO VOLTAMMETRY	10
Detection of Overlapped Peaks	10
On-Line Voltammetric Pattern Recognition	16
Deviation Pattern Recognition	19
Quantitative Voltammetric Analysis	23
Classification of Electrode Processes	25
Electrochemical Structural and Activity Classifications	29
CONCLUSIONS	36
REFERENCES	39

Accession For

NEW GRAAI	<input checked="" type="checkbox"/>
FBI TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Classification	

SEARCHED INDEXED
SERIALIZED FILED
APR 1968
FBI - NEW YORK

A

LIST OF ILLUSTRATIONS

Fig. 1.	Four steps in pattern recognition	4
Fig. 2.	Two pattern classes in a 3-dimensional feature space	6
Fig. 3.	Linear discriminant (LD) classification. W is a weight vector. F1 and F2 are feature axes	7
Fig. 4.	Doublet linear sweep voltammograms which are separated by less than 50 mV. Taken from reference 19	11
Fig. 5.	The clustering of 8 real systems in feature space. Measurements were made at several scan rates producing the spread of the patterns within each cluster. Taken from reference 19	14
Fig. 6.	The extraction of SEP features. Taken from reference 20	17
Fig. 7.	The extraction of a typical SEP derivative feature. Taken from reference 20	16
Fig. 8.	The distribution of synthetic singlets in a Fourier transform coefficient feature space. Taken from reference 19	18
Fig. 9.	Computer network for on-line pattern recognition	20
Fig. 10.	Deviation pattern obtained by fitting data which describe two closely spaced waves to single-component theory. Taken from Reference 16	22
FIG. 11.	Typical cyclic voltammograms for 4 different electrode processes. (Taken from Ref. 25).	33
Fig. 12.	On-Line voltammetric pattern classification	37

LIST OF TABLES

Table 1.	Parameter ranges for synthetic SEP doublet/singlet data base (taken from reference 18)	13
Table 2.	Electrode processes for theoretical training set	26
Table 3.	Variable levels for factorial design to study chemical structural effects on voltammetric data	33

BACKGROUND

The use of the computer has revolutionized the field of electrochemistry. Computer control of electrochemical instrumentation allows the application of an incredible range of sophisticated experiments, and provides for both the collection and analysis of data. As in IR and NMR spectroscopy, the application of computerized transform methods have been very productive also.

One area in which the computer should play an increasingly important role in the future is in the interpretation of electrochemical response and defining appropriate action. Although these functions are not likely to be taken over completely by the computer, there appear to be several areas in which artificial intelligence methods may prove to be extremely useful tools for the electrochemist.

Electrochemical systems are by nature multidimensional. A typical current-voltage response is a function of not only the identity and bulk concentration of an analyte, but also of temperature, time, solvent, electrolyte, pH, electrode composition and dozens of other variables. While humans have great difficulty in visualizing responses which are more than 3-dimensional, the computer is not limited by high dimensionality. And it is this attribute which provides significant new insight to electrochemical data.

One branch of artificial intelligence which has already been applied to problems in electroanalysis is computerized pattern recognition. Pattern recognition techniques were developed during the 1960's primarily for the solution of problems in character recognition and voice communication with computers. Isenhour, Jurs, Kowalski and Bender (1,2) introduced the concepts of pattern recognition to chemistry in the early 1970's, where most applications have been in the field of spectral interpretation. Within electrochemistry, pattern recognition has been used for the detection of severely overlapped peaks, classification of electroanalysis mechanism, classification of chemical structure, determination of herbicidal activity, and for quantitation of components producing overlapped voltammetric peaks.

Because the applications of pattern recognition to electrochemistry have now reached a significant level of achievement, and because routine applications in commercial instrumentation now appear feasible with powerful microprocessors, it is the objective to provide here a summary of achievements that have occurred over the past decade. This summary is preceded by a brief introduction to the principles of pattern recognition. More detailed background information can be obtained from several excellent texts (3-8).

PRINCIPLES OF PATTERN RECOGNITION

The ultimate goal of computerized pattern recognition is the characterization of unknowns. This is usually done in four steps: measurements, preprocessing, feature selection and classification (see Fig. 1).

In the first step, one or more transducers produce an array of measurements for each compound or chemical system to be determined. This array of measurements is called a pattern. An electrochemical example of a pattern would be a cyclic linear sweep voltammogram (an array of current measurements). Since a pattern may be made up of measurements from more than one transducer, these measurements may vary widely in magnitude and be in different units.

In the preprocessing step the measurements are scaled to normalize variations in magnitude, and mathematical transformation may be applied to produce new descriptors called features which are more useful in classifying unknowns. The transformation may be as simple as multiplying each measurement by a scale factor (weighting), or it may make features from linear combinations of the original measurements. The Karhunen-Loeve expansion is an example of one such transform (3). The Fourier transform (9) and the Hadamard transform (10) are also commonly used in the preprocessing of electrochemical data. Since an infinite set of features could be extracted from any set of measurements, the knowledge of the computer operator is important in this stage of pattern recognition. For example, if one wanted to classify cyclic voltammograms as to their reversibility, it would be advantageous to make sure that features indicative of peak separation were included.

It is desirable to have as few features as possible for computational simplicity and statistical confidence in classification (11). Thus, the third step is to eliminate as many features as possible without adversely affecting classification. The criterion for rejecting or retaining a feature might be "chemical intuition", a figure of merit such as the Bhattacharyya distance (4), or performance in actual classification. Unfortunately, the only way to be sure of picking the best subset of the original features is to try all

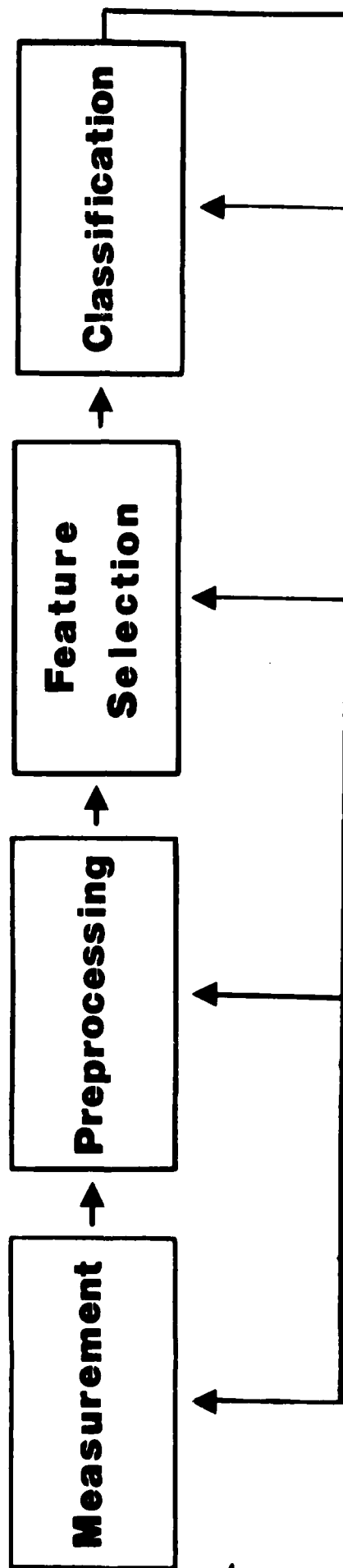


FIG. 1. Four steps in pattern recognition.

possible combinations. Even with high-speed computers, this is not often feasible from the standpoint of time, so a sub-optimal approach is most always used.

In the last step of pattern recognition, each compound or chemical system is identified by assigning it to some predefined class. The classifications are almost always made by considering the location of the patterns in a hyperspace formed by using each of the features as an axis. This is illustrated in Fig. 2. Two common classifiers which work on this principle are the linear discriminant function and the k nearest neighbor (kNN) rule.

In the case of the linear discriminant, a d -dimensional decision surface is constructed so as to define class boundaries. This surface is a line for $d=2$, a plane for $d=3$ and a hyperplane for higher dimensionalities (Fig. 3). The decision surface is defined by a weight factor, W , which is normal to the surface and passes through the origin. The position of any pattern with respect to the decision surface can be found by taking the dot product, S , of the weight vector and the vector, X , describing the pattern's position.

$$s = W \cdot X = |W| |X| \cos \theta = \sum_{i=1}^n w_i x_i + w_{n+1} \quad (1)$$

If the product is positive, the pattern lies on the same side of the surface as the weight vector; if negative, it is on the opposite side; and if zero, the pattern falls directly on the decision surface. Assigning class one arbitrarily to the positive side, and class two to the negative side, we have the following criteria for pattern classification:

$s > 0$	for class one
$s < 0$	for class two
$s = 0$	no decision

There are many methods available for positioning ("training") the decision surface (15). The Linear Learning Machine has been used extensively in chemical applications, but it is appropriate only when classes are

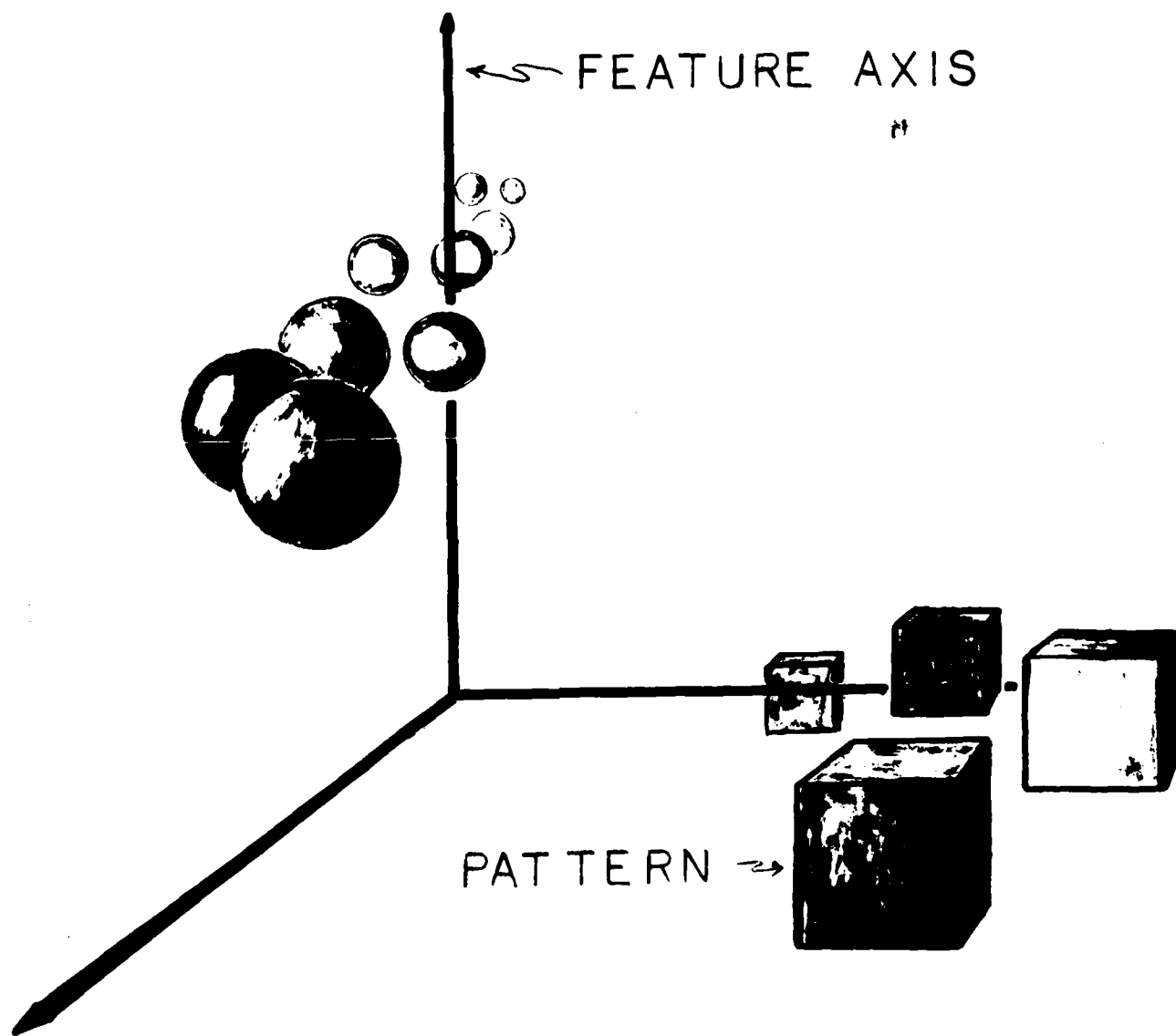


FIG. 2. Two pattern classes in a 3-dimensional feature space.

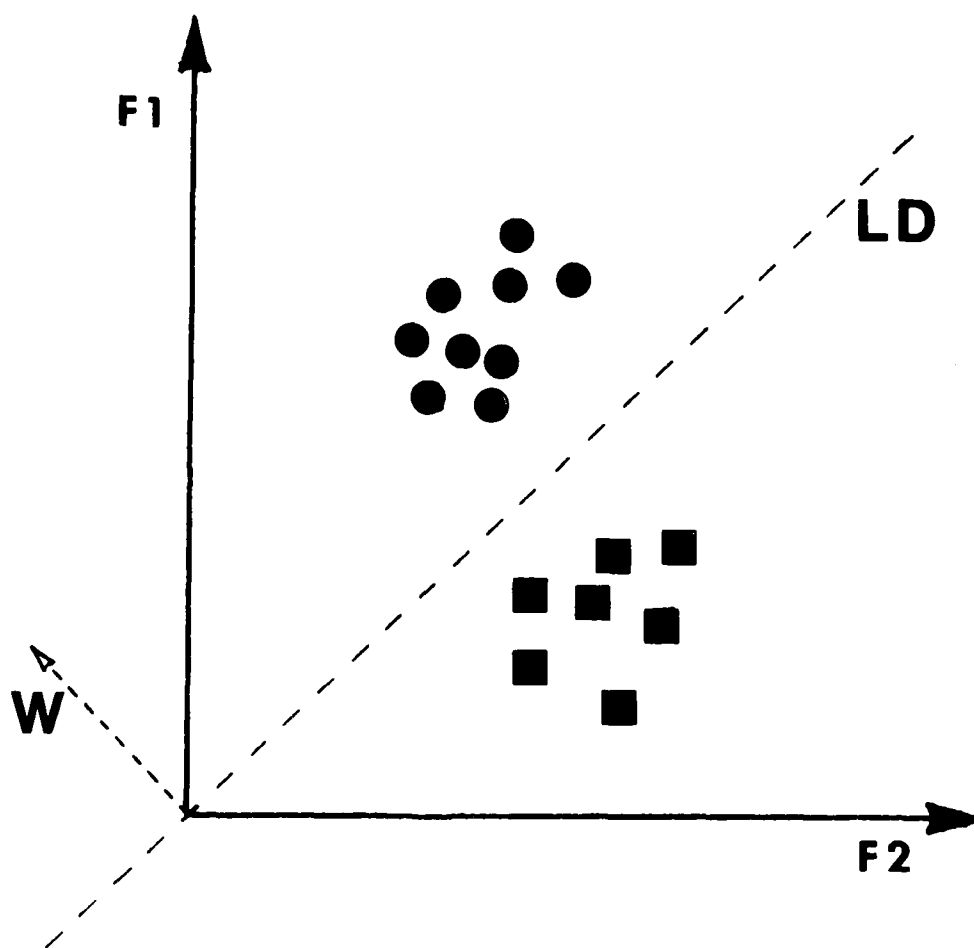


FIG. 3. Linear discriminant (LD) classification. W is a weight vector. F_1 and F_2 are feature axes.

completely linearly separable. If classes are not linearly separable, a more sophisticated training algorithm (12,13) or a different type of classifier is necessary.

The k nearest neighbor classifier assigns each pattern to the class which has the greatest representation among its k nearest neighbors. For two class problems, k is set to an odd integer to avoid the problem of tie votes. Euclidian distance, D, is usually taken as the measure of nearness. Equation 2 gives the Euclidian distance between patterns "a" and "b" in an n-dimensional feature space.

$$D_{ab} = \left[\sum_{i=1}^n (x_{ia} - x_{ib})^2 \right]^{1/2} \quad (2)$$

Other classifiers have been designed, but have not been used in electrochemical applications (5,14,15). The linear discriminant or the k nearest neighbor classifier will work well in most situations. The linear discriminant is particularly well suited to time critical decisions. Although the positioning of the discriminant may be a long process, classifications of unknowns are very rapid and require little computer memory. Nearest neighbor classifiers are better suited to multiclass problems and spaces in which the class distributions are oddly shaped. Classification is much slower, however, and a computer using this method must have a large amount of memory or have rapid access to a mass storage device.

The four steps just described are usually first carried out on a set of known patterns called a training set. The true identity of each pattern is compared to the identity assigned to it in the classification step and the results are fed back into one or more of the previous steps for error correction. This feedback process is called supervised learning and is the crux of machine intelligence.

Once the computer has learned to distinguish between different classes of patterns in the training set satisfactorily, it is tested with a different group of known patterns called a prediction set which represent the same classes included in the training set. If patterns of the prediction set are

also correctly classified, then unknown patterns may be analyzed as long as each belongs to one of the classes represented in the training set. If this cannot be assured, then the computer should be able to decide if a new pattern lies within its previous range of experience.

The specific operations which are executed in the four steps of pattern recognition are dependent on the type of data which is being analyzed and the classifications which are desired. Each step will be described more fully here in relation to several electrochemical problems.

APPLICATIONS OF PATTERN RECOGNITION TO VOLTAMMETRY

DETECTION OF OVERLAPPED PEAKS

It has been argued that one area in which automatic classification is conspicuously unsuccessful is in deciding how many narrow individual bands contribute to an observed wide, unresolved maximum on, for example, an absorbance-wavelength envelope (16). It has been pointed out, that since the problem is linear, it is always possible to choose a sufficiently large number of narrow bands which by superposition reproduce the observed maximum with any desired precision. This argument is not necessarily relevant for investigations of peak multiplicity in electroanalytical data. Although voltammetric methods are not particularly selective - i.e., overlap problems occur frequently - they are quite restrictive. That is, it is often possible to limit the number of electroactive components which can possibly contribute to a voltammetric peak by proper selection of chemical or instrumental conditions. Thus, the pattern recognition approach can become of great utility to the detection of overlapped voltammetric peaks.

Many voltammetric waveshapes, such as those obtained in linear sweep voltammetry, are broad and asymmetric. Visual detection of peak multiplicity is difficult when the separation of overlapped peaks is less than 50 mV (Fig. 4). Plotting the second derivative of the current can aid in the visual detection of two-component (doublet) voltammograms, but at separations of less than 14 mV, inspection of the second derivative gives no indication of a second component. For separations less than 14 mV pattern recognition provides the only effective technique for multiplicity classifications.

The initial work in this area used a training set consisting of Cd(II), In(III) and Sb(III) linear sweep voltammograms (17). In(III) and Cd(II) were strongly overlapped. A linear discriminant was able to detect the presence or absence of any component for a wide variety of concentrations. Classification accuracy was well above 90%.

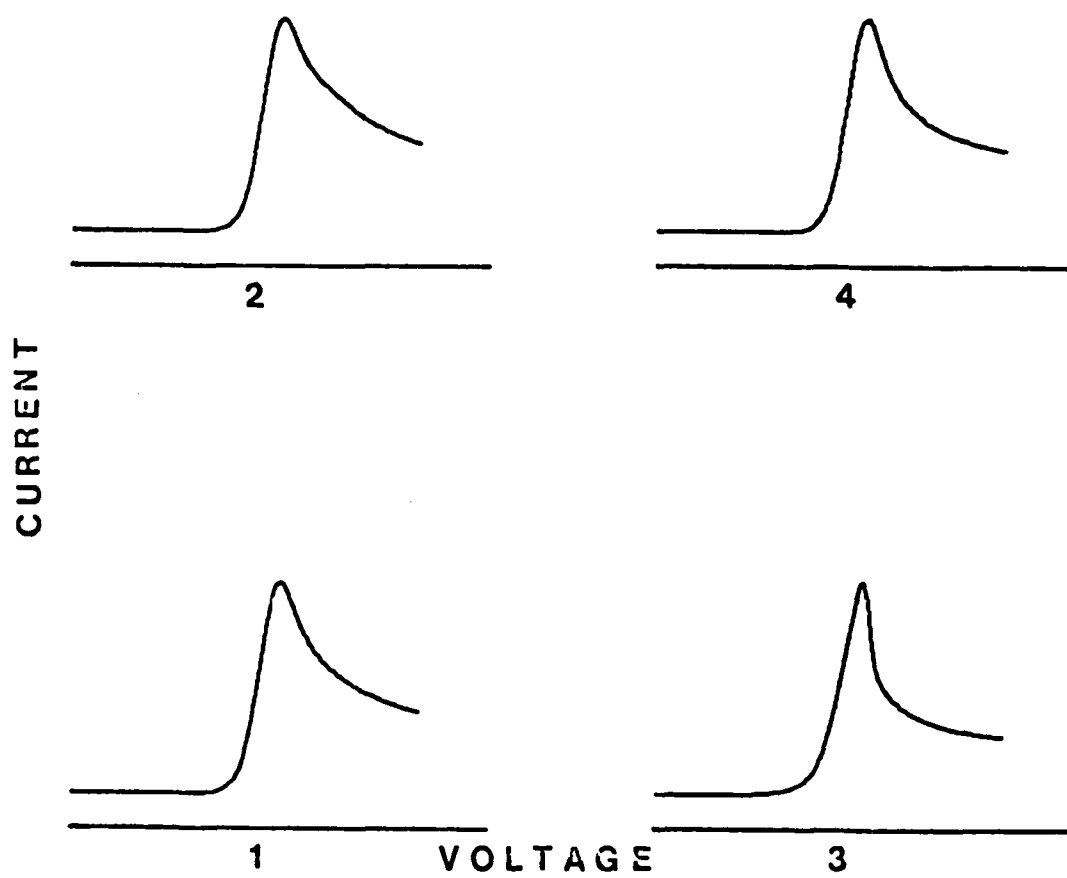


FIG. 4. Doublet linear sweep voltammograms which are separated by less than 50 mV. Taken from reference 20.

The main problem with the use of the 3-component experimental training set was that the method lack generality. Only Cd-In-Sb analysis could be done. No effort was made to extract features of an overlapped system that could be universally applied.

Theoretical training sets have been used to much advantage in generalizing singlet-doublet recognition. By creating synthetic linear sweep voltammograms with a wide range of characteristics, it has been possible to anticipate much of the real data that have been encountered. Synthetic doublet-singlet training sets have been generated for highly overlapped reversible, quasi-reversible and irreversible systems. The range of the electrochemical parameters which described the training sets is given in Table 1.

The use of synthetic data to classify real data does have some drawbacks. Voltammograms generated from theory are free from noise, chemical interferences and instrumental distortions. If this is not taken into account in the feature selection and scaling process, features may be chosen from the theoretical set which are particularly susceptible to these problems in experimental systems. This difficulty has been minimized by adding synthetic noise to the theoretical data (18), or by using training sets which contain a mixture of both real and synthetic data.

Another problem with using synthetic data to anticipate a wide range of unknown systems is that an excessive number of training set patterns may be required to define the structure of a feature space well enough for perfect classification. In singlet-doublet pattern recognition, we discovered that real patterns sometimes fell into empty regions of feature space even when the training set contained several thousand synthetic patterns. Again, the inclusion of real data in the training set was helpful (19). The wide range of synthetic data in the training set allowed the classification of totally new voltammograms, and the real training set voltammograms increased the classification accuracy for patterns which fell within the fairly restricted regions of feature space where real data is normally found (Fig. 5).

Two types of features have been used in doublet-singlet recognition. The first type is based on the symmetry of the voltametric waveform or its derivatives. Typical features of this type are shown in Fig. 6.

TABLE 1. Parameter ranges for synthetic SEP doublet/singlet data base (taken from reference 18).

ψ	.01, .02, .05, .10, .20, .50, 1.0, 2.0, 5.0, 10.0
α	.3, .4, .5, .6, .7
n^a	.9, 1.0, 1.1; 1.8, ..., 2.2; 2.7, ..., 3.3
$Ep_2 - Ep_1$	4, 6, 8, 10 mV
C_1^*/C_2^{*b}	10:1 to 1:10
a)	... represents increments of .1
b)	$C_2^* = 0$ for singlets, all other values chosen at random

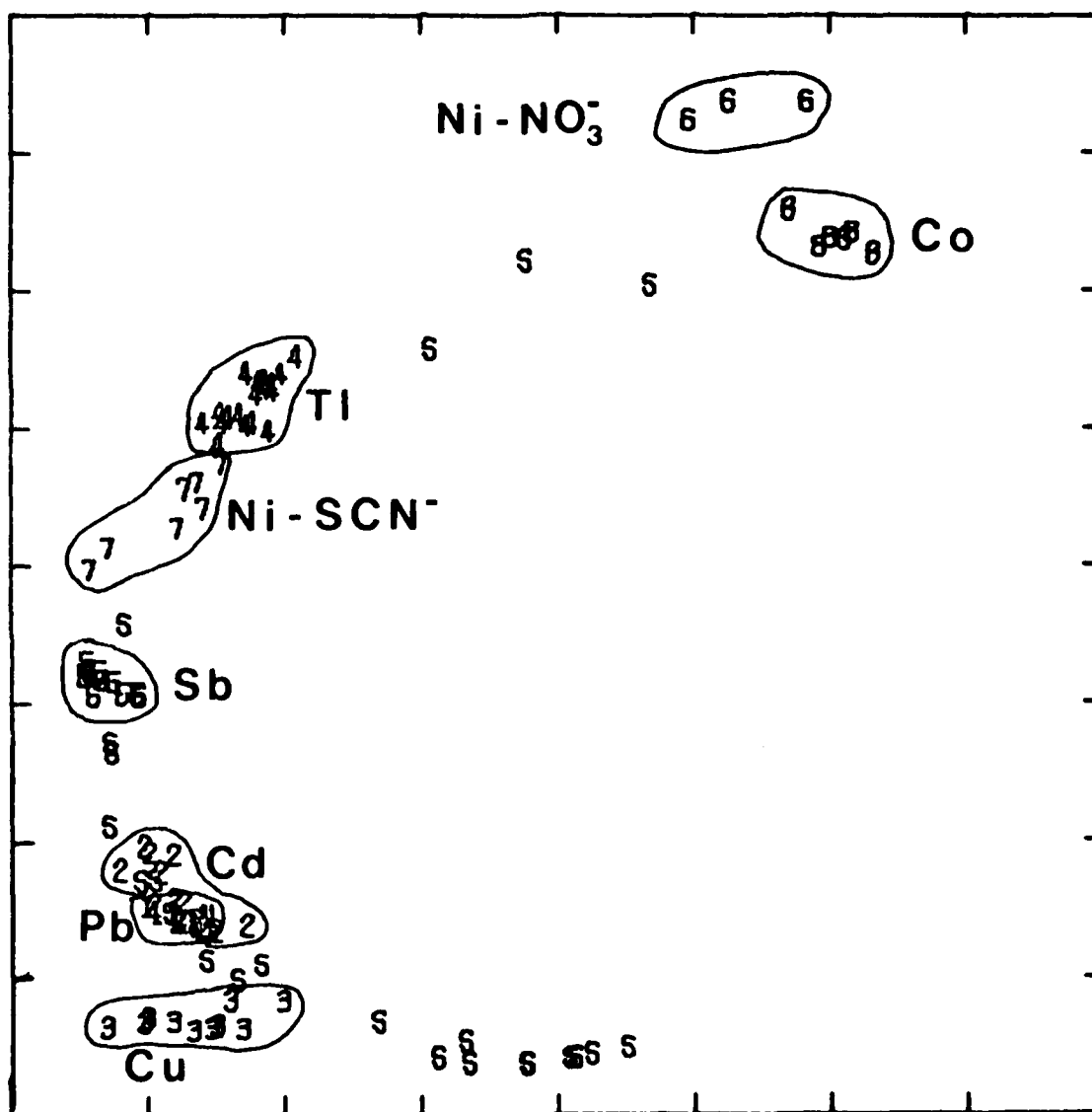
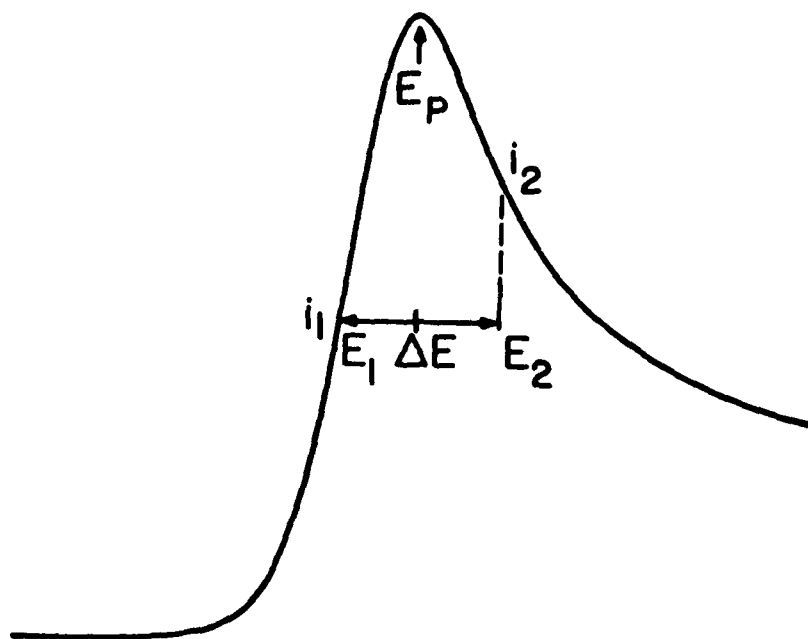


FIG. 5. The clustering of 8 real systems in feature space. Measurements were made at several scan rates and concentrations producing the spread of the patterns within each cluster. S's correspond to theoretical singlets in the same feature space. Taken from reference 19.



Typical Features

$\Delta i = i_2 - i_1$ at $\Delta E = 8, 16, 24, 32$ mV about E_p

$$I_{\text{function}} = i_2 * i_1 / (i_2 - i_1)^2$$

FIG. 6. The extraction of SEP features.

One hundred and thirty-three (133) such features have been defined and are referred to collectively as the SEP feature set (18).

An alternative set of features for describing the shape of a voltammogram may be obtained through the use of the Fast Fourier Transform (FFT) (Fig. 7). The low frequency components of the transformed data contain most of the fundamental shape information, while the higher order coefficients represent noise information and may be disregarded (18).

Of the two feature sets used, the SEP features have the advantage in classification of patterns outside the limits of the training set. Fourier transform coefficients on the other hand, gave consistently higher classification accuracies for data within the training set limits, indicating that they are more sensitive to minute differences in shape such as are seen between singlet and highly overlapped doublet voltammograms.

The kNN classifier performed better than the linear discriminant in the recognition of singlets and doublets in the synthetic training sets. This is due to the fact that the singlet class distributions are oddly shaped and cannot be easily partitioned from the doublets with a linear discriminant. Figure 8 shows a plot of a group of synthetic singlets as they are distributed by two Fourier transform features that were useful in classification. The doublets, which have not been plotted, are encircled by the singlet curve.

ON-LINE VOLTAMMETRIC PATTERN RECOGNITION

Pattern recognition of voltammetric data has the greatest value when applied in an on-line fashion. Rapid classifications allow the optimization of measurements. Instrumental distortions can be detected and corrected before an entire set of bad data is recorded, and unanticipated electrochemical behavior can be recognized and investigated. In our laboratory, single-doublet pattern recognition was the first application to be implemented on-line, so that voltammetric assays could be constantly monitored for unexpected highly overlapped peaks. Subsequently, all of the voltammetric pattern recognition studies in our laboratory were conducted on-line.

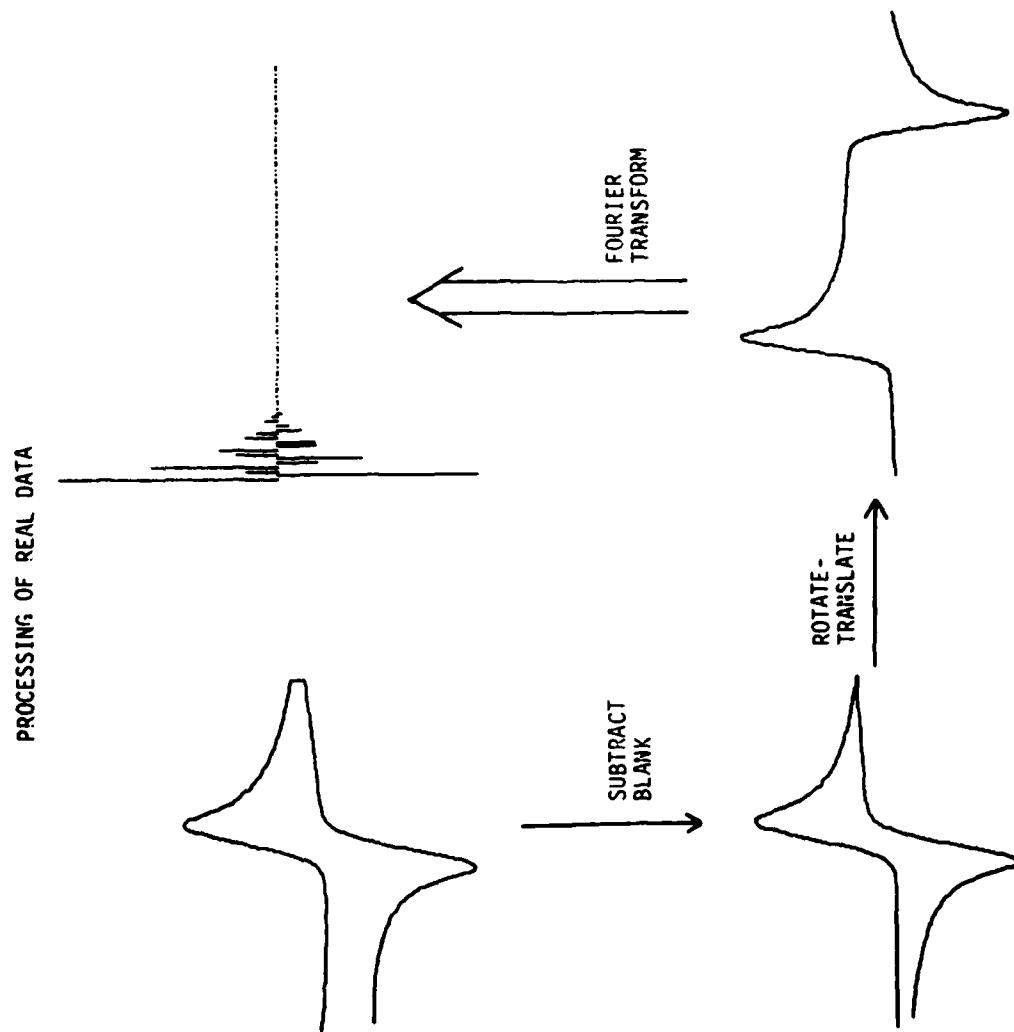


FIG. 7. The extraction of fourier transform features for cyclic voltammetry.

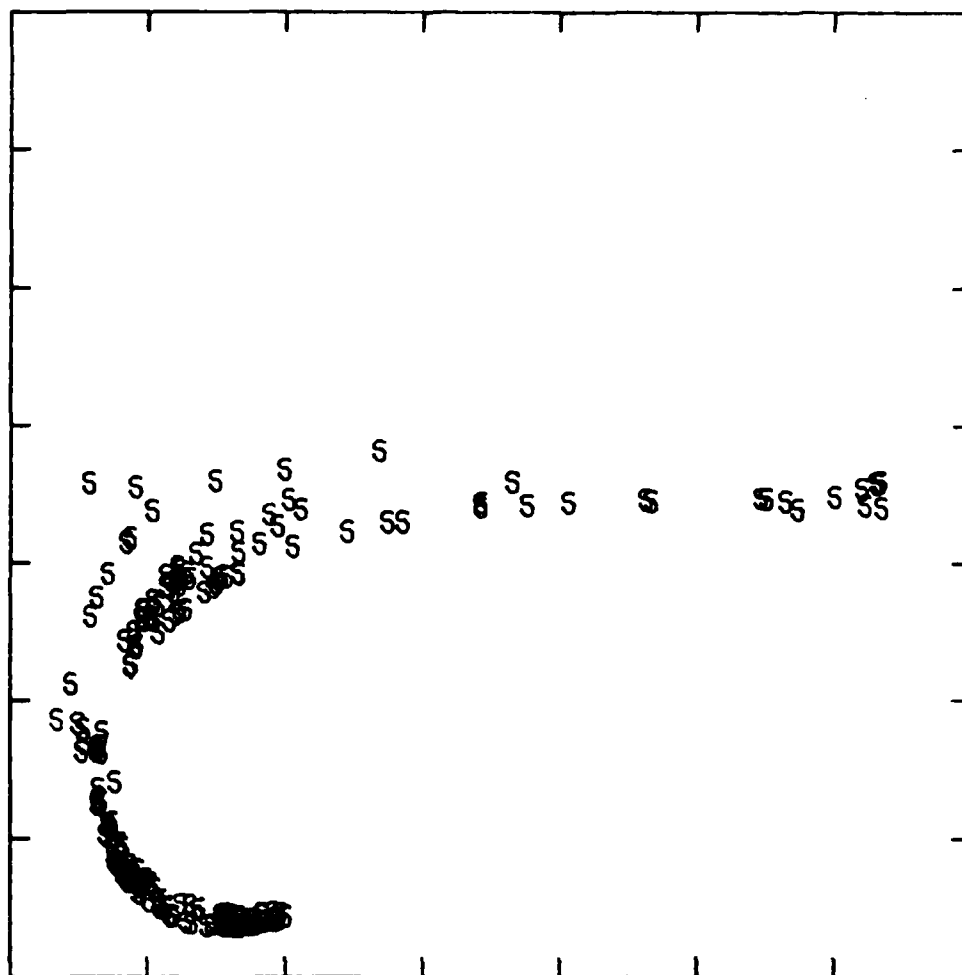


FIG. 8. The distribution of synthetic singlets in a Fourier transform coefficient feature space. Taken from reference 20.

We have found computer networking to be very helpful when doing pattern recognition on-line. In our laboratory, three Hewlett-Packard mini-computers and two microcomputers were available for electrochemical experimentation. The fastest minicomputer was equipped with a 5 Mbyte disk drive, a printer, plotter and other peripherals, and was interfaced to the other mini's via high-speed, bidirectional, 16-bit paralleled interfaces (Fig. 9). The microcomputers could be linked to any of the three minicomputers with RS-232 serial interfaces. This network allowed the tasks of data collection and analysis to be delegated efficiently. The fast, well-equipped computer performed the pattern recognition calculations while the other computers were used for experimental control. Using this network, a voltammogram could be transferred from an experimental control computer to the pattern recognition computer, stored on disk, analyzed for multiplicity, and the results returned to the control computer in 12 seconds (20). Other analyses (structural, activity, mechanism) could be conducted with similar response time.

DEVIATION PATTERN RECOGNITION

Hypotheses concerning the mechanisms of electrode reactions can often be formulated in terms of equations relating a measurable quantity such as a cell current to an independent variable - e.g., time or applied potential. Provided that the equation adequately accounts for the experimental data, kinetic parameters can be derived by fitting the model equation to the data by regression. The chemist can subsequently apply statistical methods such as the F-test to judge the validity of a proposed mechanism.

Although statistical tests do not guarantee that a proposed mechanism is correct, they do provide a means of rejecting an incorrect hypothesis. Unfortunately, once a test has resulted in rejection, no guidance to a better understanding of the experimental evidence is provided.

Meites and coworkers have introduced the "deviation plot"/"deviation pattern" concepts to overcome this difficulty (16,21,22). In any experiment, there may be systematic as well as random errors; moreover, in the interpretation of an experiment, systematic errors may be introduced by assuming a model mechanism which happens to be wrong. It is often possible

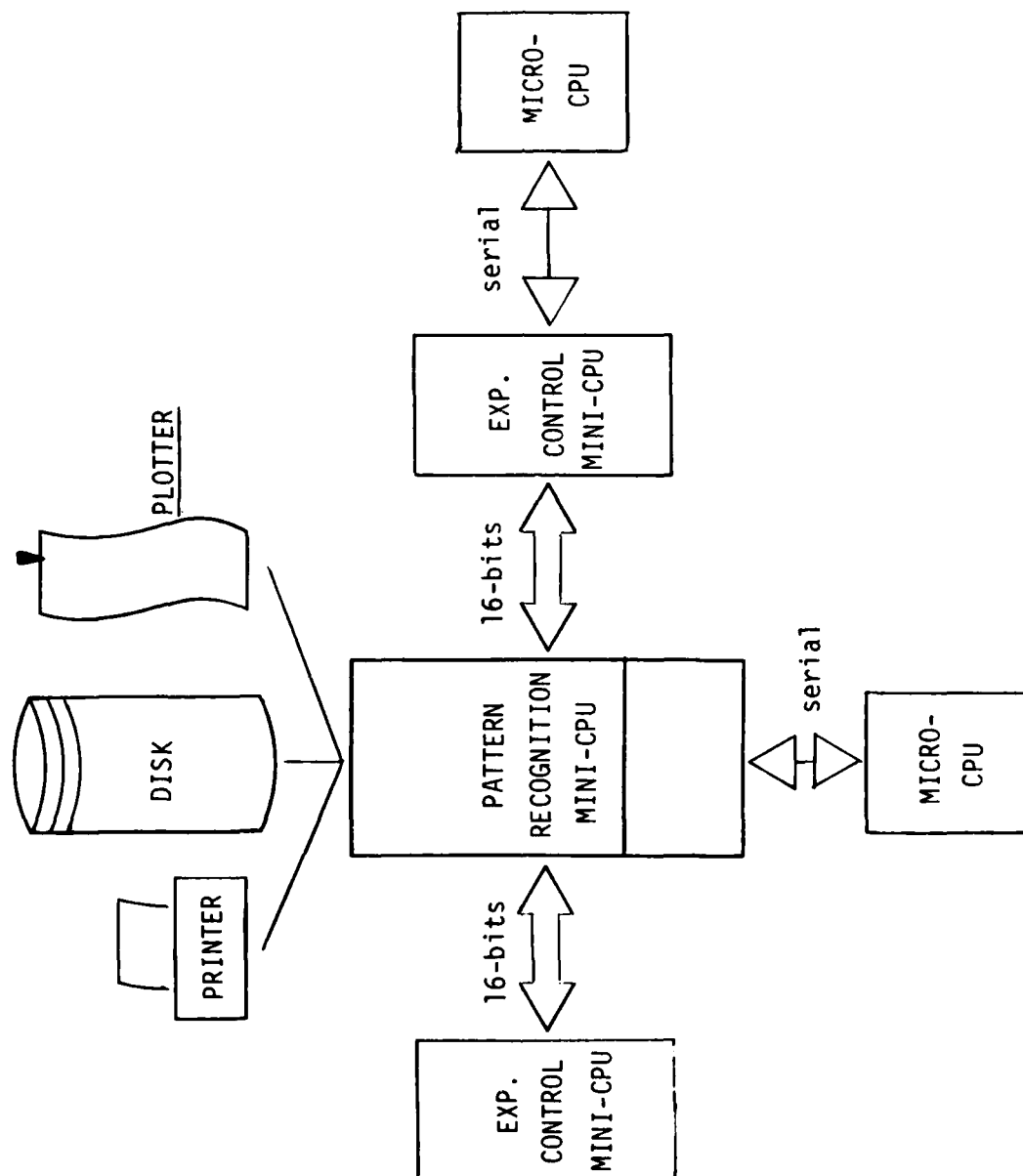


FIG. 9. Computer network for on-line pattern recognition.

to eliminate or correct for systematic experimental errors. In such cases, the features of a "deviation plot" may assist the experimenter not only to decide whether his hypothesis is wrong, but also give him guidance to select a more probable theory.

Corresponding to a given hypothesis, a number, m , of experimental data points, $y_{m,meas}$, are used to obtain a fit of the model assumed, and the best values of the parameters - rate constants, half-wave potentials, etc. - are calculated. Subsequently, the dependent variable, $y_{m,calc}$, of the experiment (e.g., the current measured in a polarographic experiment) is calculated based on the hypothesis and the parameters thus determined. The deviation plot is then obtained by plotting, as a function of the independent variable, the quantity: $\Delta y_m = (y_{m,meas} - y_{m,calc})/S$, where S is the standard deviation from regression analysis.

If data, free from variable systematic errors, have been used to obtain a fit of a model that describes them correctly and completely, the deviation plot will consist of a set of points randomly scattered around the horizontal axis. If, however, the equation used for its construction is wrong, the deviation plot will exhibit some systematic variation (a "deviation pattern") superimposed on the random noise. In such cases, the general features of the plot may guide the chemist to select a better hypothesis, if a library of "deviation patterns" is available to him. Such a library should contain entries giving the (noise free) theoretical "deviation patterns" obtained by fitting a model to theoretical data calculated by assuming a different model. To the electrochemist, such a library greatly facilitates the work involved in selecting and testing hypotheses. Figure 10, taken from reference (16), illustrates a sample deviation pattern obtained by fitting a model, describing a single polarographic wave, to theoretical polarographic data corresponding to two merging polarographic waves.

If the chemist has assumed that a polarographic wave originates from one electroactive substance, he can test this hypothesis by construction of the corresponding deviation plot. If the plot exhibits a systematic variation, he can reject the hypothesis. Moreover, if this variation resembles that of the deviation pattern of Fig. 10, there is evidence that he should test the

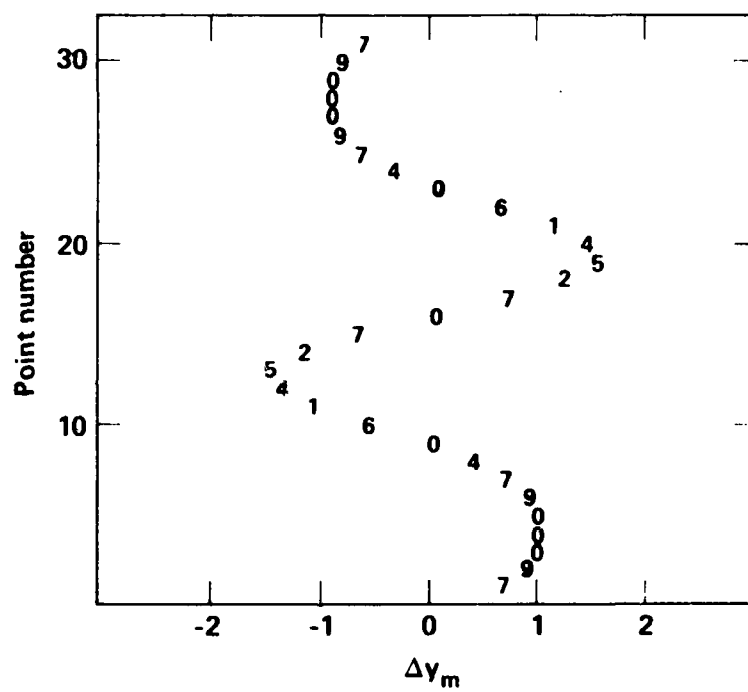


FIG. 10. Deviation pattern obtained by fitting data which describe two closely spaced waves to single-component theory.

hypothesis that two electroactive species are present. These tests and the recognition of deviation patterns can be carried out by a digital computer (16,21), which attempts to assign a given deviation plot to an entry in the library of deviation patterns. Once the scientist has assembled his own library for his field of interest, he can rapidly decrease the number of possibilities that have to be further investigated for the interpretation of data from new experiments (16). The limitations of the approach are, of course, that no information is provided on any hypothesis not represented in the library - and, like the F-test, the method cannot give any guarantee that a hypothesis is correct.

QUANTITATIVE VOLTAMMETRIC ANALYSIS

Pattern recognition produces discrete classifications, so it is not by nature a quantitative technique. There are situations, however, when the only information desired is whether or not an unknown is within a desired range of concentrations. Pattern recognition may be used for rapid determinations of this type. All patterns in the training set which are within the desired concentration range can be assigned to one class, and the rest of the patterns to another, allowing classifiers to be trained by traditional methods. More exactly, quantitative estimates may be obtained by considering the distance of a pattern from a decision surface if proper training procedures are employed. For instance, a linear discriminant can be trained to recognize if a component is contained within a mixture. If a mixture pattern falls on the side of the discriminant which indicates that a particular component is present, the training can also be such that the dot product of the weight vector and the mixture pattern gives the concentration of that component.

Bos and Jasink (23) have used linear discriminants to facilitate the determination of cadmium, lead and thallium down to 10^{-8} M by anodic stripping voltammetry at a hanging mercury drop electrode. The three metals were preconcentrated in the mercury electrode by electrolysis at -900 mV vs. S.C.E. in a non-complexing medium and subsequently stripped off the electrode by an anodic scan (differential pulse mode). Voltammograms obtained in this manner exhibit severely overlapped peaks due to the stripping of thallium and

lead and a less pronounced overlap of the thallium and the cadmium peaks (E_{peak} approximately -380, -445 and -580 mV vs. S.C.E. for lead, thallium and cadmium, respectively). Thus, large concentrations of thallium may prevent an accurate determination of lead and cadmium by conventional methods.

To obtain quantitative results using the pattern recognition approach, experimental single-component anodic stripping voltammograms, taken at seven different concentrations, were used to train three linear discriminants, one for the detection and quantification of each component. The discriminants were positioned by a least-squares algorithm, which minimized the error in concentration predictions. The heavy metal concentrations were predicted quite accurately for selected solution compositions in the concentration range 10^{-6} - 10^{-8} M. Even in such cases, where the peak due to 5×10^{-8} M lead completely disappears under the peak due to 10^{-6} M thallium, the predictions of all three metal concentrations were quite accurate (10-20%). Although the study ignored the possibility of selectively shifting the reduction potentials of lead and cadmium sufficiently negative to avoid interference, the principles of the approach could prove useful to the electrochemist.

Ichise et al. (24) have used a similar procedure to quantify lead(II), thallium(I), cadmium(II) and zinc(II) in staircase voltammetry. An "analog feedback" learning machine approach was used to position the linear discriminant. The computer algorithm of the analog linear learning machine is very simple: Contrary to the digital linear learning machine, it has no decision part, and it is trained by the feedback of a simulated analog output from an adaption part. The price paid for this simplicity is, however, an increased time necessary to predict the composition of an unknown sample. The approach has been shown to produce rather accurate predictions for samples containing lead(II), thallium(I), cadmium(II) and zinc(II) at the 10^{-3} M concentration levels. For the practicing analyst, who is often interested in the determination of these elements at the nano- to micromolar levels, the approach could be valuable if applied to stripping analysis. At present, it seems very probable that if the efforts involved in developing pattern recognition methods are accompanied by efforts towards optimizing the experimental conditions, a useful approach to quantitative electrochemical analysis may evolve.

CLASSIFICATION OF ELECTRODE PROCESSES

One of the most exciting uses of pattern recognition in electrochemistry has been in the examination of electrode processes. Many processes can occur at the electrode-solution interface besides the simple transfer of electrons. Electroactive species may be adsorbed, form amalgams, or undergo a great variety of chemical reactions before or after electron transfer. These complications may have distinct effects on the shape of an electrochemical waveform which can be used by a pattern recognition analysis to characterize the mechanism of the electrode reaction.

Cyclic linear sweep voltammetry can generate a large amount of fundamental mechanistic information. The reverse sweep allows the examination of products generated in the forward sweep, and can detect processes which might be occurring to remove these products from the solution. Many cases have been treated theoretically allowing the generation of a broad synthetic training set. For these reasons it has been used by us to provide a data base for mechanistic pattern recognition. The electrode processes which have been examined by pattern recognition are listed in Table 2 (25). Figure 11 illustrates the kind of subtle variations in wave shape associated with different mechanisms.

The features which are traditionally used in the determination of mechanism with cyclic voltammetry are the separation of the cathodic and anodic peaks, the ratio of the peak currents, and variation of the peak separation and current ratio with scan rate. Comparison of these features to those expected from theory will in many cases confirm or reject a proposed mechanism, and for this reason these were the first features considered for pattern recognition analysis. However, there are some cases in which these features are so similar for several different plausible electrode processes that no decision can be made. The applicability of pattern recognition in such cases has been tested (25) using a training set which contained 885 synthetic patterns representing all of the electrode processes listed in Table 2. Since previous studies indicated that Fourier transform features were very sensitive descriptors of curve shape, they were included in the training process along with the more traditional features.

TABLE 2. Electrode processes for theoretical training set (taken from reference 25).

	<u>Abbreviation</u>	<u>Process</u>
<u>Electrode Phenomena:</u>		
Uncomplicated Reversible	Uncomp	$O + ne \xrightleftharpoons{\quad} R$
Quasireversible	QR	$O + ne \xrightleftharpoons[k_b]{k_f} R$
Amalgam Formation	Amalg	$O + ne \xrightleftharpoons{\quad} R(Hg)$
Weak Adsorption	Ads	$O_{ads} + ne \xrightleftharpoons{\quad} R$
<u>Coupled Chemical Reactions:</u>		
Preceding	CE	$Z \xrightleftharpoons{K} O$
		$O + ne \xrightleftharpoons{\quad} R$
Reversible Following	EC_{rev}	$O + ne \xrightleftharpoons{\quad} R$
		$R \xrightleftharpoons{K} Z$
Irreversible Following	EC_{irr}	$O + ne \xrightleftharpoons{\quad} R$
		$R \xrightarrow{k_f} Z$
Catalytic Following	EC_{cat}	$O + ne \xrightleftharpoons{\quad} R$
		$R \xrightarrow{k_f} O$

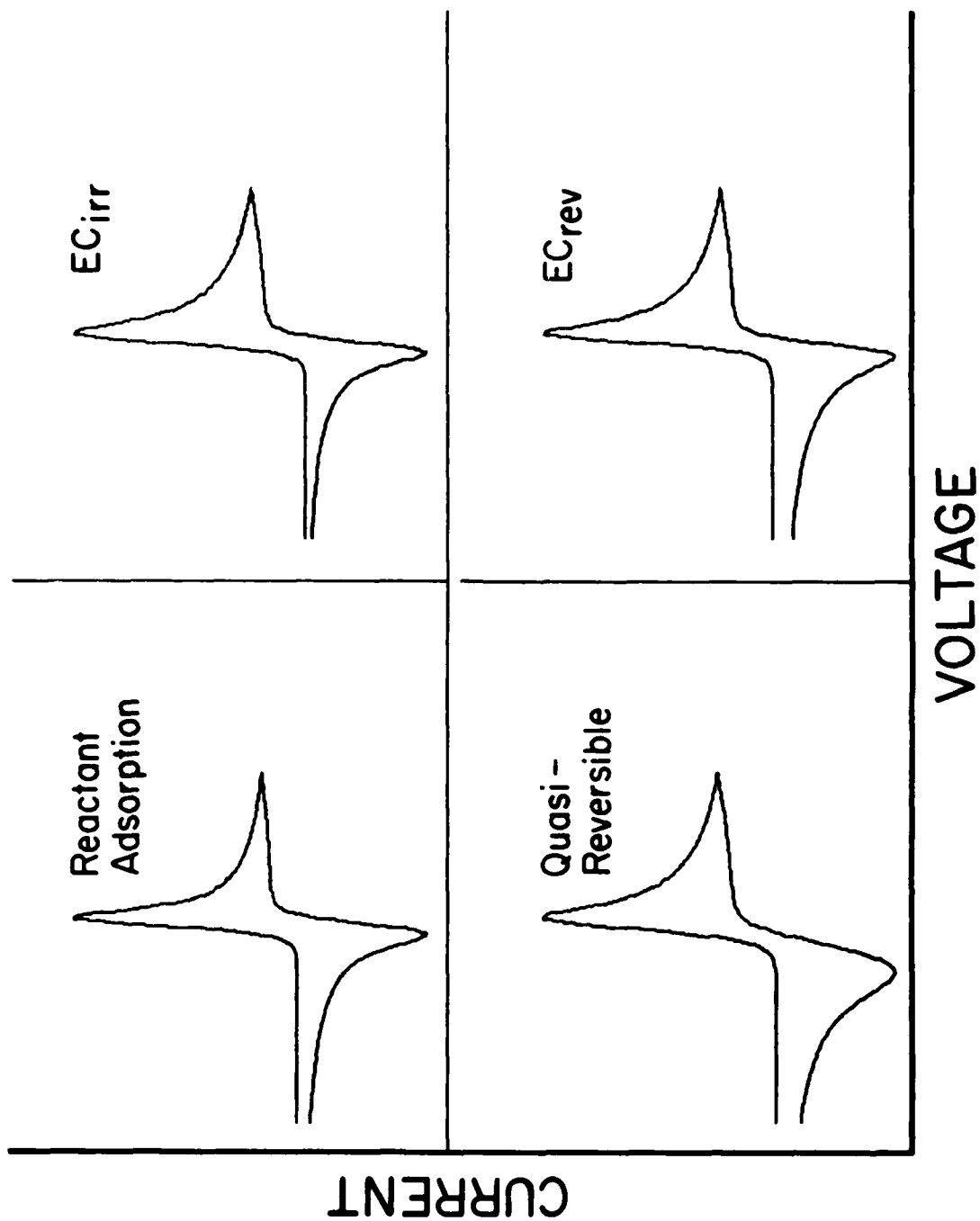


FIG. 11. Typical cyclic voltammograms for 4 different electrode processes.
(Taken from Ref. 25).

Through the use of an iterative feature selection procedure and the kNN classifier the computer learned to classify training set patterns with 96.7% accuracy. Twelve experimental systems were then run at a variety of scan rates to test predictive accuracy. Five systems were classified correctly every time while most of the rest had classification accuracies above 90% (25).

Perhaps the most significant aspect of this work, however, is that only a single voltammogram is required to achieve accurate classification. This should be contrasted with the need to examine behavior trends over many experiments for conventional mechanistic studies. Nevertheless, the single-voltammogram classification procedure should be used primarily as a quick diagnostic, to be followed by more detailed study. When viewed in this perspective single-voltammogram classification becomes a useful tool for screening and categorizing large numbers of samples.

The pattern recognition analysis had several interesting implications. One surprise was that none of the traditional features calculated from peak separations and current ratios were retained in the feature selection routine. This suggests that although these parameters are easily calculated, their exclusive use in interpretation of cyclic linear sweep voltammograms is not justified. The necessity of using a computer precludes the consideration of discrete Fourier transform coefficients in all applications, but it might be possible to define some set of easily calculated features which are more sensitive to electrochemical mechanism than peak separations and current ratios.

Another finding was that experimental cyclic staircase voltammograms could be used with the theoretical linear sweep training set for classification of the electrode process. Although the similarity between linear sweep voltammograms and staircase voltammograms obtained using a sampling parameter of 0.7 has been noted before, the successful classification of the staircase data shows that the two techniques are nearly identical for a variety of electrode processes. The staircase technique is superior experimentally since the double layer charging current can be minimized, but it has not often been used due to the small number of systems which have been examined theoretically.

An added advantage in using pattern recognition techniques based on large theoretical data sets is that semiquantitative estimates of the parameters describing the electrode process can be obtained as by-products of the kNN classification process. When an unknown pattern is matched to a synthetic pattern, the parameters that were used to generate the synthetic pattern can be assigned to the unknown. DePalma and Perone (26) have used this technique in a number of uncomplicated systems to determine the number of electrons transferred; ψ , a measure of the irreversibility; and α , the transfer coefficient.

ELECTROCHEMICAL STRUCTURAL AND ACTIVITY CLASSIFICATIONS

The classification of chemical structure using electrochemical techniques, is a challenging problem. Voltammetric responses lack fine structure and probably will never compete with IR or NMR spectroscopy in qualitative analysis. The complex dependence of an electrochemical response on many variables, and theoretical problems in relating structure to electrochemical activity, make qualitative voltammetric analysis even more formidable.

Even though the difficulties in qualitative electroanalysis are great, the rewards of developing a reliable means of structural identification through electroanalysis would also be great. Due to recently developed miniaturization techniques, electrodes are the most promising probes of in vivo chemical species. Carbon fiber electrodes may be implanted within a single cell or neuron (27). Electrochemical detectors in liquid chromatography are becoming very important because of their high sensitivity and selectivity. Quantities of electroactive material in the picogram range have been analyzed. Osteryoung, et al. (28) have recently demonstrated the feasibility of scanning the potential of a liquid chromatographic electrochemical detector, so the development of qualitative voltammetric methods would open up the possibility of the characterization of eluants that are 1000 times less concentrated than those which can be analyzed by spectroscopic techniques.

Linear-free-energy relationships have generally been the most useful expressions for relating structure to electrochemical activity in the past. A substituent group will have a characteristic effect on the free energy of an electrochemical reaction occurring in its vicinity. This effect may occur through electron withdrawal, electron donation, or it may be steric in nature. In any case, the effect may be quantified through the use of Hammett substituent constants. For a given class of electrochemical reactions, there will be a linear relationship between $E_{1/2}$ and the substituent constants σ (29).

There are two main problems in the use of linear-free-energy relationships. The first and largest problem is the determination of the reaction series to which an unknown belongs. Such a deduction from electrochemical behavior is not straightforward. Furthermore, there may be several reaction series which may be constructed for a class of compounds depending on solution conditions. The slope of the $E_{1/2}$ vs σ plot would be different at high pH's due to a change in the mechanism of reduction.

The second main problem is that there is often not enough $E_{1/2}$ separation for different substituents or substituent combinations to allow for confidence in identification, especially when experimental reproducibility is low due to uncontrolled matrix effects. The consideration of more information than $E_{1/2}$ would clearly be helpful.

Because pattern recognition is well suited to the consideration of large amounts of information and to making use of obscure relations, we have applied it to chemical structure identification from electrochemical data. The main questions have been what information should be collected and how much?

Burgard and Perone (30), used staircase voltammetry to analyze 29 compounds belonging to four different electroactive group/skeleton combinations. The classes examined were aromatic-nitro, aliphatic-nitro, aromatic-aldehyde and aromatic-aliphatic-ketone. These classes were almost completely separated on the basis of peak potential, but since this feature alone cannot be considered sufficient for many identification problems, the voltammograms were examined for any shape information which might characterize a particular electroactive group or the skeleton to which it was attached. It was found that the change in peak shape with scan rate produced fair

classifications (70% correct), but that complete separation of the classes was not possible for the experimental conditions and compounds which were chosen. The results suggested that the information content of the electrochemical data base should be increased for more reliable structural classifications.

Ichise, Yamagishi and Kojima (31-33) have proposed the simultaneous determination of complete E-i-c and C_{dl} -E-c patterns (c = surface concentration) and have published several papers on instrumentation and data compression algorithms for reaching that goal. E-i-c patterns were generated by applying a pseudo-random waveform to the cell and monitoring the current response. The surface concentration of the depolarizer was calculated from the current in an analog fashion with an " $s^{-1/2}$ module" which eliminated the effect of diffusion. C_{dl} was obtained by applying a high frequency 10 mV sinusoidal wave to the cell and measuring the amplitude of the 90 degrees out-of-phase component of the current.

The idea of obtaining double-layer capacity information may be fruitful. The capacitance of the double layer is dependent on adsorption of the analyte, and the strength and potential dependence of adsorption may indicate the presence of certain functional groups (34). π -electron interaction between adsorbed molecules and the electrode surface has a characteristic influence on the adsorption behavior of organic substances (34), and specific interactions between the analyte and some other molecule or ion within the double layer may also be helpful in identification (35,36). Some adsorbed organics will inhibit the reduction of metal ions, while others, through the so called "cap-pair" effect will accelerate reductions (37).

The use of a potential-step technique such as cyclic staircase voltammetry represents a simple alternative to Ichise's method (32) of obtaining information on both adsorption and electron transfer kinetics. The current decay immediately after a step is primarily capacitive while current at later times is almost totally due to electron transfer reactions. Thus, measuring the current at several times during each step and by changing the scan rate, information on both the kinetics of the electrode process and the differential capacity can be obtained with a single sweep.

As is true with cyclic linear sweep voltammetry, the reversal of the scan is important in detecting chemical reactions which succeed the electron

transfer step. Immediate repetition of a cyclic scan will detect any products which may have been generated in the reverse scan of the first cycle.

One additional parameter which can be explored is the "drop hang time". This refers to the time period between the creation of a stationary mercury drop and the beginning of the first staircase scan. During the waiting time, a potential can be applied. This variable was investigated in our work to see if there was any class specific information in the kinetics of adsorption.

Another source of structural information is the electrochemical response of the analyte to chemical perturbations. Changes in solution conditions have been useful in classical studies of structure-activity relationships. Exploration of a variety of solutions will help define the best conditions for particular classification problems.

All of the experimental and solution variables which have been examined systematically in our classification studies are listed in Table 3. The determination of the effect of each of the seven variables is difficult without good experimental design. To characterize all main effects and all interactions one could arrange the experiments by a factorial design (38). For the seven variables considered here, 128 runs would be needed for each compound. The large number of runs can be avoided by using a saturated fractional factorial design (39) in which the main effect of all seven variables can be investigated in only eight experiments. By running a second fraction, in which all variable levels have been reversed from their state in the first fraction, all confounding between the main effect of variables and the interaction of two variables will be eliminated. Higher order interactions (the interaction of three or more variables) may still be confounded with the main effects, but in most cases such interactions are relatively small in magnitude.

In our work (40), a fractional factorial design was used in which all of the main variable effects have been de-aliased as described above. In addition, one of the experiments run early in the analysis of each compound is repeated near the end of the analysis to determine instrumental precision and to detect any decomposition of the sample. This makes a total of 17 voltammograms which must be taken for each compound. These experiments yield 17 current-voltage and 17 differential capacity curves for each compound.

TABLE 3. Variable levels for factorial design to study structural effects on voltammetric data.

VARIABLE NUMBER	VARIABLE	LOW LEVEL (-)	HIGH LEVEL (+)
X ₁	% Ethanol	0.5 %	9.5 %
X ₂	pH	8.0	5.1
X ₃	Surfactant Concentration	0	1.4 x 10 ⁻⁵ M
X ₄	Number of Cycles	1	2
X ₅	Scan Rate	0.25 V/s.	1.0 V/s.
X ₆	Drop Hang Time	0.2 s.	30 s.
X ₇	Sampling Time	30% of step ($\alpha'=.7$)	end of step ($\alpha'=.007$)

Graphical analysis of the error involved in the calculation of variable effects was done for several nitroaromatics and nitrodiphenyl ethers (40). It was discovered that all of the variables chosen for study had significant effects on the Faradaic responses of the compounds examined. The magnitudes of the effects and the shapes of the effect curves were quite different, indicating that redundant information was not recorded. All of the variables also had a significant effect on the differential capacity curves of strongly adsorbed species, but some of the effects could not be distinguished from noise for more weakly adsorbed compounds. Only pH, number of cycles and % ethanol had a significant effect on the capacitance response of both weakly and strongly adsorbed organics.

Since the variables chosen and the levels over which they were changed seemed to be appropriate for most compounds from a signal-to-noise perspective, the variable effects were further examined for any information which might be useful in structural classifications. Forty-five compounds representing three major structural classes were chosen, and features derived from the variable effects were tested for predictive ability (41). Class 1 consisted of 19 nitroaromatics containing a single benzene ring; Class 2 contained nine nitrodiphenylethers, and Class 3 consisted of 17 azo compounds. The classes were completely overlapped in potential, and all compounds were reduced by the same number of electrons, so the identification of the classes from their voltammetric behavior was not a trivial problem.

The pattern recognition analysis revealed that all of the variables produced structural-specific information. Most of the information was found in the Faradaic responses. Changes in the Faradaic responses with the number of cycles gave the highest classification accuracy of 93.3%. Scan rate changes yielded 89%, while pH, surfactant and drop hang time all produced classification accuracies of 84%. Changes in a Faradaic response with % ethanol and sampling time appeared to contain the least structural information, giving classification accuracies of 66.7 and 75.6%, respectively. As was expected from the signal-to-noise analysis, the effects of the several variables on the capacitive responses were much poorer structural predictors. Classification accuracies ranged between 60.0 and 75.6%.

Although changes in differential capacity responses caused by changes in the experimental variables were not very helpful, the shapes of differential capacity curves which were obtained under the same experimental conditions were excellent structural descriptors. Using shape features derived from differential capacity curves taken under one set of experimental conditions, 93.3% classification accuracy was achieved. Four other sets of experimental conditions yielded over 90% classification accuracy.

An interesting sidelight of the organic structural classification study was that herbicidal activity could also be predicted (41). The nitrodiphenylethers could be divided into compounds which were strong herbicides and those compounds which showed little or no herbicidal activity. Both Faradaic and capacitive responses could be used to separate these classes for over half the experimental conditions examined.

The ability of voltammetric responses to predict the herbicidal activity can be explained by the mechanism of herbicidal action for the nitrodiphenylethers. It is thought that these compounds are involved in the initiation of destructive free radical reactions with the phospholipid molecules which make up cellular membranes (42). Since the first step in the reduction of aromatics at the mercury electrode also involves the formation of radical species (43), some similarity between herbicidal activity and voltammetric behavior can be expected.

CONCLUSIONS

It is clear that, at this point, the feasibility and performance of the methods which combine pattern recognition and electrochemistry (PR/EC) have been established. The present interpretive capability exceeds considerably what could be achieved by more conventional methods. Computerized identification of severely overlapped peaks, characterization of heterogeneous kinetics or mechanism of the electrode process, and qualitative identification of analytes by structural type, on the basis of interpretation of a single voltammogram and conducted rapidly by an on-line laboratory computer, are examples of unique achievements of PR/EC. Figure 12 summarizes this current capability. The important question now is in what directions future work will go.

There are some obvious future developments or extensions of the capabilities already demonstrated. Applications of PR/EC to qualitative identification of liquid chromatographic effluents using scanning electrochemical detection should be very fruitful. Still another area of application would be in vivo monitoring of biological fluids with PR/EC.

In addition to the above extensions, it would appear worthwhile to investigate the applicability of pattern recognition to other electrochemical measurements. For example, we have already shown that battery lifetime prediction could be achieved by using pattern recognition and cluster analysis methods to interpret charge/discharge data from new Ni/Cd batteries (44). It would be worthwhile to investigate the applicability of pattern recognition to the interpretation of transient battery test methods with respect to establishing state-of-charge, predicting imminent failure, identification of failure mode, lifetime prediction, etc.

Another intriguing area of investigation encompasses "cybernetic instrumentation" -- i.e., laboratory instrumentation aided by artificial intelligence (A.I.). The goal would be to utilize the interpretive powers of PR/EC to develop adaptive control functions for voltammetric instrumentation. Of course, these same principles can be applied subsequently to other instrumental systems, but we have already demonstrated that voltammetric

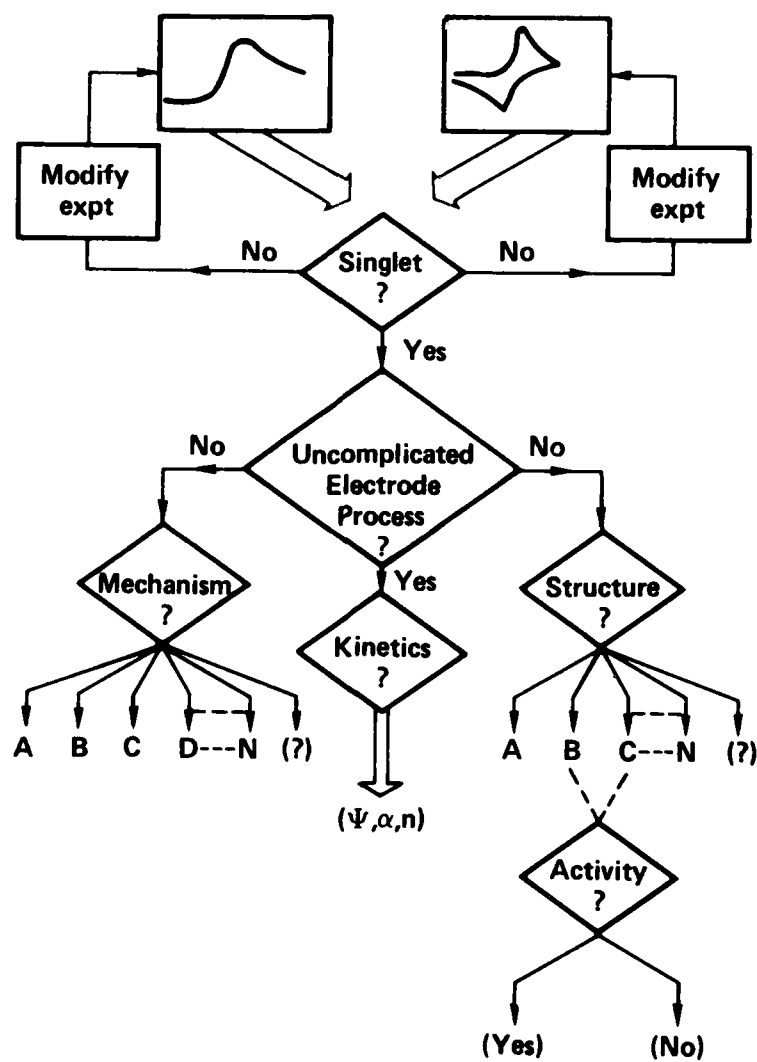


FIG. 12. On-Line voltammetric pattern classification.

techniques provide a favorable proving ground for on-line pattern recognition studies. Because we have also developed the capability for sophisticated computer control and feedback in voltammetric experimentation (45,46), it would appear feasible to develop AI-guided electrochemical instrumentation.

This work supported by the Office of Naval Research (Purdue University) and the U.S. Department of Energy Contract W-7405-Eng-48 Lawrence Livermore National Laboratory.

REFERENCES

1. T. L. Isenhour and P. C. Jurs, Analytical Chemistry, 43, 20A (1971).
2. B. R. Kowalski and C. F. Bender, J. Amer. Chem. Soc., 94, 5632 (1972).
3. J. M. Mendel and K. S. Fu, "Adaptive Learning and Pattern Recognition Systems"; Academic Press: New York, p.56-60 (1970).
4. K. Fukunaga, "Introduction to Statistical Pattern Recognition"; Academic Press: New York, pp. 258-285 (1972).
5. B. G. Batchelor, "Practical Approach to Pattern Classification"; Plenum Press: London (1974).
6. P. C. Jurs and T. L. Isenhour, "Chemical Applications of Pattern Recognition", Wiley-Interscience, New York (1975).
7. H. C. Andrews, "Mathematical Techniques in Pattern Recognition"; Wiley-Interscience, New York (1972).
8. K. Varmuza, "Pattern Recognition in Chemistry", Lecture Notes in Chemistry, Vol. 21, Springer-Verlag, New York (1980).
9. L. E. Wanger, N. M. Frew, T. L. Isenhour and P. C. Jurs, Appl. Spectrosc., 25, 203 (1971).
10. B. R. Kowalski and C. F. Bender, Anal. Chem., 45, 2234 (1973).
11. J. W. Sammon, D. Foley and A. Proctor, Proceedings of the 1970 IEEE Symposium on Adaptive Processes, University of Texas, p. IX.2.1 (1970).
12. J. Schechter and P. C. Jurs, Appl. Spectrosc., 27, 225 (1973).
13. G. R. Brissey, R. B. Spencer and C. L. Wilkins, Anal. Chem., 51, 2295 (1979).
14. S. Wold, Pattern Recognition, 8, 127 (1976).
15. B. R. Kowalski, Anal. Chem., 52, 112R (1980).
16. L. Meites, CRC Critical Reviews in Analytical Chemistry, 8, 1-53, (1979).
17. L. B. Sybrant and S. P. Perone, Anal. Chem., 43, 382 (1971).
18. Q. V. Thomas and S. P. Perone, Anal. Chem., 49, 1369 (1977).
19. Q. V. Thomas, R. A. DePalma and S. P. Perone, Anal. Chem., 49, 1376 (1977).

20. R. A. DePalma and S. P. Perone, Anal. Chem., 51, 825 (1979).
21. L. Meites and G. A. Shia, "Automatic Elucidation of Reaction Mechanisms in Stirred-Pool Controlled-Potential Chronocoulometry", in "Chemometrics: Theory and Application", Bruce R. Kowalski, Editor, ACS Symposium Series, 52, 127-152 (1977).
22. L. Lampugnani, L. Meites, P. Papoff and T. Rotunno, Annali di Chimica (Rome), 65, 257-266 (1975).
23. M. Bos and G. Jasink, Anal. Chim. Acta, 103, 151 (1978).
24. M. Ichise, H. Yamagishi and T. Kojima, J. Electroanal. Chem., 113, 41 (1980).
25. S. D. Schachterle and S. P. Perone, Anal. Chem., 53, 1672 (1981).
26. R. A. DePalma and S. P. Perone, Anal. Chem., 51, 829 (1979).
27. J. L. Ponchon, R. Cespuglio, F. Gonon, M. Jouvot and J. F. Pugol, Anal. Chem., 51, 1483 (1979).
28. R. Samuelsson, J. O'Dea and J. Osteryoung, Anal. Chem., 52, 2215 (1980).
29. P. Zuman, "The Elucidation of Organic Electrode Processes"; Academic Press: New York, 1969, Chapter 2.
30. D. Burgard and S. P. Perone, Anal. Chem., 50, 1366 (1978).
31. M. Ichise, H. Yamagishi and T. Kojima, J. Electroanal. Chem., 94, 187 (1978).
32. M. Ichise, H. Yamagishi, H. Oishi and T. Kojima, J. Electroanal. Chem., 106, 35 (1980).
33. M. Ichise, H. Yamagishi, H. Oishi and T. Kojima, J. Electroanal. Chem., 108, 213 (1980).
34. B. B. Damaskin, O. A. Petrii and V. V. Balrakov, "Adsorption of Organic Compounds on Electrodes"; Plenum Press: New York, 39-40 (1971).
35. S. Gupta and S. Sharma, Electrochim. Acta, 10, 151 (1965).
36. E. Dutkiewicz and A. Puacz, J. Electroanal. Chem., 100, 947 (1979).
37. K. Sykut, G. Dalmata, B. Nowicka and J. Saba, J. Electroanal. Chem., 90, 299 (1978).
38. C. D. Hendrix, CHEMTEC, 9, 167 (1979).

39. G. E. P. Box, W. G. Hunter and J. S. Hunter, "Statistics for Experimenters"; John Wiley and Sons: New York, 1978; Chapter 12.
40. W. A. Byers and S. P. Perone, Anal. Chem., In Press.
41. W. A. Byers and S. P. Perone, Anal. Chem., In Press.
42. G. Orr, Ph.D. Thesis, Purdue University, W. Lafayette, IN (1981).
43. B. Kastening and L. Holleck, J. Electroanal. Chem., 27, 355 (1970).
44. W. A. Byers and S. P. Perone, J. Electrochem. Soc., 126, 720 (1979).
45. Q. V. Thomas, L. Kryger and S. P. Perone, Anal. Chem., 48, 761 (1976).
46. S. P. Perone, D. O. Jones and W. F. Gutknecht, Anal. Chem., 41, 1154 (1969).

END

FILMED

1-83

DTIC